DOCUMENT RESUME

ED 074 144 TM 002 509

| | |
|---|---|
| AUTHOR | Sewell, Alan F.; Dornseif, Allan W. |
| TITLE | Controlled Multivariate Evaluation of Open and Traditional Education at the Junior High School Level. Preliminary Report. |
| SPONS AGENCY | Spencer Foundation, Chicago, Ill. |
| PUB DATE | Feb 73 |
| NOTE | 13p.; Paper presented at annual meeting of the American Educational Research Association (New Orleans, Louisiana, February 25-March 1, 1973) |
| EDRS PRICE | MF-$0.65 HC-$3.29 |
| DESCRIPTORS | Academic Achievement; Grade 7; Grade 8; *Open Education; *Program Evaluation; Social Development; Speeches; Student Adjustment; Student Attitudes; Technical Reports; *Traditional Schools |

ABSTRACT

A year-long study to evaluate the relative educational outcomes of open and traditional education is being conducted at the O. W. Huth Upper Grade Center, Matteson, Illinois. Midpoint analyses and evaluations of the study are presented. The open plan group includes 140 randomly assigned 7th- and 8th- graders in a single, specially constructed classroom. An equal number of randomly assigned control students pursue the traditional departmental program. The open plan teachers follow an interdisciplinary curriculum. Pretest, midpoint, and posttest measures are presented in four areas: academic achievement, personal growth, social development, and attitudes. (Author)

ED 074144

Preliminary Report

# Controlled Multivariate Evaluation of Open and ~~the~~ Traditional Education at the Junior High School Level

Alan F. Sewell
Department of Psychology
DePaul University
Chicago, Illinois

and

Allan W. Dornseif
School District 162
Matteson, Illinois

TM 002 509

This study is being funded by a grant provided by the Spencer Foundation, Chicago, Illinois. The authors also wish specifically to acknowledge the assistance and cooperation of Mr. Donald Stevenson, Principal, O. W. Huth Upper Grade Center and Dr. Mari J. K. Brown, DePaul University. Mr. W. R. Jarosz, DePaul University, aided greatly in data processing, and Miss K. Billingham, DePaul University, aided in all phases of administration, data collection, and scoring.

ASSISTANT PROFESSOR,
DEPARTMENT OF PSYCHOLOGY

ALAN F. SEWELL, PH.D.
DIRECTOR OF THE OFFICE OF
SPONSORED PROGRAMS AND PROJECTS

DEPAUL UNIVERSITY
2323 N. SEMINARY AVE.
CHICAGO, ILLINOIS 60614

549-6900
EXT. 683

This report provides preliminary (midpoint) results of a year-long educational evaluation project conducted at the O.W. Huth Upper Grade Center, School District 162, Matteson, Illinois. A final report will be prepared shortly after the conclusion of the 1972-73 academic year. Rather than to present meaningful statistical results, the purpose of the present report is to describe the educational program which is still being evaluated, to establish the rationale for the evaluative measures employed, to describe those measures, and to note various problems which have developed in the course of the project.

## The School District

Elementary School District 162 includes all or parts of five suburban communities at the southern end of Cook County, Illinois. These five communities are Matteson, Olympia Fields, Park Forest, Richton Park, and Park Forest South. The area includes some long-established settlements and a large number of newer housing developments and apartment complexes.

The socioeconomic characteristics of the district are quite diverse. One community within the District was ranked fourth in "class status" in the Chicago Metropolitan Area last year; another area within the district contains housing supported by FHA 235 legislation and open only to families having more than three children and earning less than $8,000 per year. Inhabitants of other portions of the district are the families of upward-bound junior executives of large national corporations, skilled blue-collar workers, and assemblers employed at local factories.

Approximately 3,400 children attend the district's eight schools. Of these, some 2,600 children attend the seven schools which extend from kindergarden through sixth grade, and 800 attend O. W. Huth Upper Grade Center (grades seven and eight). While the district as a whole is predominantly white, two of the schools are enrolling increasing percentages of black students; enrollment at another school includes a small proportion of children of oriental ancestry; still larger members of Mexican-American children attend another school. Children attend local, neighborhood schools through sixth grade; hence the full heterogeneity of the district's population is not expressed anywhere other than in the Upper Grade Center.

## Development of the OSCAR Program

The educational program of the Upper Grade Center had long been of the traditional variety. An awareness of new and changing needs led, in the past few years, to the establishment of additional programs in introductory home economics, industrial arts, and guidance counseling. During this period of program review and revision, the Upper Grade Center was also experiencing an obvious need for expansion of physical facilities. The Upper Grade

Center building had been designed for occupancy by 600 students but had an actual enrollment of more than 800. After an attempt to place the school on a year-round basis failed, a bond issue to remodel and enlarge the building was passed.

An "open classroom" was planned and designed as part of the total remodeling and enlargement program. A space equivalent to that of five traditional classrooms and their adjacent hallway was allocated to the "open classroom": a single, undivided area approximately 140 feet long and 40 feet wide. Plans for the development of a "school within a school" to be housed within this area were undertaken by district personnel under the leadership of the principal of the Upper Grade Center.

(The fortuitous acronym OSCAR was soon applied to the open classroom project. This acronym memorialized the given name of the former superintendent of the district, for whom the school was named, and simultaneously represented "Open Space for Conceptualizing Attitudes and Responsibilities.")

The OSCAR program, as developed, called for staffing by four teachers and two teacher-aides for a total of 140 seventh and eighth grade boys and girls to be assigned to the open classroom. One teacher was chosen from each of four discipline areas: language arts, mathematics, science, and social studies. These teachers were selected on the basis of expressed interest in and probable capability for the OSCAR program. During the summer of 1972 they attended a ten-day, multi-talent development training program which emphasized the discovery and development of talents unique to each student. The OSCAR concept required these teachers to plan and work together in content areas and to develop a multi-disciplinary approach to subject matter.

During the Spring of 1972, two groups of children were randomly chosen: an OSCAR group and a control group, 140 children in each group. Half of each group was chosen from the school's existing enrollment of then-seventh graders, and half was chosen from the district's population of then-sixth graders (who would attend the Upper Grade Center as seventh graders during the 1972-73 academic year). The random assignment procedures were designed so as to ensure representation of the entire school district. Males and females were equally represented within the assignment groups. Distribution of sex and grade level within the two groups is summarized in Table 1.

Table 1
Composition of the assigned study population. Each group includes proportional representation of the district's "sending schools."

| | | OSCAR | Control |
|---|---|---|---|
| 7th Grade | Girls | 35 | 35 |
| | Boys | 35 | 35 |
| 8th Grade | Girls | 35 | 35 |
| | Boys | 35 | 35 |
| Total | | 140 | 140 |

OSCAR students were informed of their assignment during the Spring of 1972. Only two of these had to be replaced because of parental objections.

Students assigned to the control group were never so informed, but in an effort to control for Hawthorne effect, the control group was assembled and designated "Mr. Stevenson's Special Group" by the school's principal, Donald Stevenson.

OSCAR: Operational Phase

Construction delays prevented occupancy of the OSCAR classroom at the opening of the 1972-73 school year. In order to maintain the OSCAR program and group, the school cafeteria was called into service as a substitute. For more than two months the OSCAR program was conducted in this overly large and acoustically poor room. Even when the OSCAR group was able to occupy the completed room, shipping problems caused delays in furniture deliveries; the expected equipment was not completely on site until mid-February, 1973. (Tables of various sizes are used instead of student desks. Students' personal belongings are stored in lockers installed along one wall of the room. The room is completely carpeted, and the ceiling is acoustically tiled.)

For various reasons, OSCAR teachers found it necessary to modify their plans with respect to multi-talent development and multi-discipline planning. The day is organized around the four major disciplines in the morning, with the amount of time devoted to each determined cooperatively by the OSCAR team. Projects and modified contract assignments are the basic activities of students in all subjects except mathematics, in which an individualized skill development materials kit is used. A considerable degree of flexibility is permitted OSCAR students: upon agreement between student and teacher, a student may opt to spend more or less time in any day on given subjects.

During the afternoon, or about one-third of the school day, OSCAR students are scheduled into other school programs, such as physical education, home economics, industrial arts, music and art. Only 75% of the OSCAR group is out of the room during this time, however. This remaining 25% stay in the OSCAR room to complete individual assignments, work in small study groups, or receive individual or small group tutoring.

Control group students, meanwhile, pursue the "traditional" departmentalized curricula and schedules of the Upper Grade Center, changing rooms and classmates at the end of each class period. The control group students are never together as a group except when they are deliberately assembled for purposes of testing. No attempt has been made to encourage interdisciplinary teaching nor the introduction of any unusual teaching techniques among non-OSCAR teachers. For purposes of this study, however, four non-OSCAR teachers, one in each of the four primary disciplinary areas, were

selected as "control" teachers without their knowledge; this was done to permit some comparisons of teacher behaviors and attitudes.

Comment

Even a cursory review of the literature reveals that "open education" and "open classroom" are highly ambiguous terms. In comparison to Neill's "Summerhill" concept of open education, the OSCAR program more closely resembles traditional education. Indeed, the freedom given the OSCAR student is fairly stringently limited; it is, nevertheless, substantially greater than that accorded the control student. The OSCAR program was developed to incorporate as many of the general characteristics of "open education" as seemed appropriate to the needs and abilities of students, teachers, and parents. Whether the OSCAR program is indeed "open education" and whether this study is a genuine evaluation of "open education" are undoubtedly fit questions for debate.

## Method

Planning for this evaluative study was a collateral of all stages of development of the OSCAR program. The meager evaluative literature of open education provided no clear recommendation for the present study, and the theoretical and conceptual literature seemed no more helpful. While the general evaluative design of the study was easily contrived, a major problem was that of conceptualizing and defining the dependent variables. Discussions between the investigators, district personnel, and the OSCAR teachers suggested various ways in which OSCAR outcomes could or should differ from non-OSCAR outcomes. An early agreement was reached to focus this study primarily upon the student, in contradistinction to other studies which have focussed primarily upon the teacher. A second area of agreement was that the educational process involves not only students and teachers, but parents as well.

The scope of this study was finally limited to a reasonable number of conceptual dependent variables which could logically be related to expected differences in the two educational approaches which functioned as independent variables. For the student populations these would be measures of academic achievement, personal growth, social development, and attitudes. For the teachers, measures of classroom behavior and attitudes would be employed. And for parents, a measure of attitudes would be used. Each of these measures is described in some detail below.

### Student: Academic Achievement

Since learning in the traditional academic areas apparently

persists as one objective of education, academic achievement was chosen as one conceptual area in which differences between OSCAR and control students might be found. The eight tests of the Stanford Achievement battery were chosen as accepted and objective measures of academic accomplishment. Scores were available for OSCAR and control subjects as of the end of the 1971-72 academic year, conveniently providing a baseline for each group. The study's design required that the same tests be administered in January, 1973 and again in May, 1973.

Student: Personal Growth

"Personal growth" can, obviously, be defined in various ways. Because accepted, objective, and standardized measures were preferably to be utilized, a number of published tests were considered. The Bell Adjustment Inventory was chosen, somewhat reluctantly, as a reasonable embodiment of the concept of "personal growth." Among the advantages of this test were: (1) a long history of its use, with attendant literature attesting to its standardization, reliability, and validity; (2) the inclusion of a large number of items which might subsequently prove useful for item analysis; (3) the existence of a set of standardized scales having face validity designations of clear relevance (e.g., "Home Adjustment"); (4) a relatively short recommended time limit. One clear and one probable disadvantage were noted at the time of this test's selection; (1) published norms were inapplicable to the populations of this study; (2) some of the items would probably be inappropriate to the age group of this study.

The study design called for this test to be administered at or near the beginning of the 1972-73 academic year, again in January, 1973, and a third time in May, 1973. The first administration actually occurred in October, 1972; booklets and answer sheets were distributed to the OSCAR students in a morning session and to the control students during the afternoon of the same day. Disadvantages of the test were noted at both sessions. Average time for completion of the 200 items of the test was greater than one hour, causing serious disruption of school schedules; further, it appeared that many of the students experienced difficulty with the reading level of the test, and many of the students required explanations of difficult or unfamiliar terms. It was necessary to continue the testing session on another day for those students unable to complete the test within available time.

The same test was readministered in January, 1973. Two new strategies were employed at this testing session. Test items were administered by a tape recording previously prepared by the senior investigator. Students were given only the answer sheet upon which they were to record their responses. Items were repeated as requested by the students, but no assistance was given in the form of explanations. This strategy permitted completion of the entire test in approximately one hour and eliminated problems associated with reading level difficulties.

The problem of misunderstood words was handled by a second strategy. All items of the test were reviewed by the senior investigator and an assistant. A list of potentially difficult or unfamiliar terms was thereby derived. (Examples of the former are "naive" and "gullible"; examples of the latter are "saps" and "blues.") A multiple-choice vocabulary test was then developed for the 25 terms thus identified. This test was administered to students of the OSCAR group and the control group. If a term was incorrectly identified on the vocabularly test by more than 20% of the students, the Bell item containing that term was deleted from the test. Since not more than two items were thereby deleted from any one scale of the test, the balance of the scales was not significantly disturbed, particularly inasmuch as published norms were not to be used.

Student: Social Development

"Social development" is also rather an ambiguous term. For the purposes of this study, a compound operational definition was employed. Certain of the scales of the Bell Adjustment Inventory (e.g., "Masculinity-Feminity," actually an interest and behavior scale) appear relevant, as do certain portions of the attitude scale to be discussed below. Similarly, portions of a self-concept scale to be administered at the final testing session are relevant to the concept of social development.

Student: Attitudes

A 24-item attitude scale was specifically developed for this study. The scale superficially employs a Likert-type technique but was actually designed to measure attitudes in twelve conceptually discrete areas according to the schema shown in Table 2.

Table 2
Design of the attitude scale. Each of the 12 attitude categories is represented by a positive and a negative item.

| | Academic | Social | Personal |
|---|---|---|---|
| Education | EA$^{+}_{-}$ | ES$^{+}_{-}$ | EP$^{+}_{-}$ |
| School District | SA$^{+}_{-}$ | SS$^{+}_{-}$ | SP$^{+}_{-}$ |
| Teachers | TA$^{+}_{-}$ | TS$^{+}_{-}$ | TP$^{+}_{-}$ |
| Self | PA$^{+}_{-}$ | PS$^{+}_{-}$ | PP$^{+}_{-}$ |

An example of a PP item is: "I am satisfied with myself as a person" (PP+). If a student recorded strong agreement with this statement, his response would be scored 5; a strong disagreement response would be scored 1. The negative version of this item would be scored in the opposite direction. The student's

total score for each attitude category, then, could range between 2 and 10.

It should be noted that data provided by this instrument would be pertinent to inferences concerning several dimensions of the study. PP data, for example, can be expected to reflect the student's self-concept, while PS data would, hopefully, reflect the student's social development.

The attitude scale was first administered to students in the OSCAR and control groups in October, 1972. A second administration occurred in January, 1973. A final administration is scheduled for May, 1973.

Variants of this attitude scale are employed with two other respondent groups and are discussed below.

Teachers: Classroom Behavior

An observational analysis of the teachers' classroom behavior was chosen as a means of securing relevant data about expected differences in instructional methodology. Flanders' method of interaction analysis was selected as being generally accepted, widely employed, and presumably, reliable. An assistant unacquainted with any of the OSCAR or control teachers was instructed in Flanders' method by the junior investigator until close agreement in results was reported by the two independent observers. Thereafter this assistant observed each OSCAR and each control teacher for the prescribed period on the same day. The first observation was conducted in November, 1972, and the second in February, 1973. The ten standard Flanders categories were employed in recording observations.

Teachers: Attitudes

An attitude scale very similar to the student's attitude scale was developed for discerning teachers' attitudes. The same attitude categories were employed; indeed, the same items were used, with the obviously necessary modifications of language and object. For example, whereas the student's PP+ item states "I am satisfied with myself as a person," the teacher's PP+ item states "I am satisfied with my students' personal development."

This scale was first administered to OSCAR and control teachers in October 1972. The second administration occurred in January, 1973. A final administration is scheduled for May, 1973.

Parents: Attitudes

A third version of the attitude scale was prepared for administration to the parents of students in both groups, again with appropriate modifications. The parents' version of the PP+ item states "I am satisfied with my child as a person."

Copies of the attitude scale were mailed to all parents of OSCAR and control students in August, 1972. Envelopes were prepared for return mailing directly to the senior investigator. A covering letter was enclosed to solicit parents' cooperation and asking the two parents to respond independently.

A second mailing occurred in January, 1973, and a third is scheduled for May, 1973.

## Preliminary Results

The general analytic design of this study called for the use of three independent variables: a grouping variable (OSCAR vs. Control), a grade level variable (seventh vs. eighth), and a sex variable (male vs. female). With respect to the major instruments employed in this study (attitude scale, Bell Adjustment inventory, Stanford Achievement Battery), the possible existence of a fourth independent variable had to be considered: subdivisions of each instrument, such as subscales of the Bell Adjustment Inventory. The dependent variable of each analysis would, of course, be the scores attained on each such subtest. The fundamental analytical model for the study is shown in Table 1 (page 2). This design obviously lends itself to analysis of variance techniques.

Prior to actual analysis, all available scores for a given instrument were utilized to compute a matrix of intercorrelations. (For example, the six scales of the Bell Adjustment Inventory were used to calculate a matrix consisting of fifteen correlation coefficients.) Each such matrix was then studied to determine whether analysis of variance or multivariate analysis of variance was the more appropriate analytic technique. The pattern of intercorrelations of the attitude scale was such as to recommend an ANOVA approach, while MANOVA was judged more appropriate for analysis of data derived from the Bell Adjustment Inventory and the Stanford Achievement Battery.

In some cases, as explained below, change scores were used, while in other cases raw scores recorded at each testing session were used in analysis. The decision whether to employ change scores or raw scores in any given analysis depended upon demands imposed by the rationale of the study, the rational of the OSCAR program, or the investigators' expectations.

Student: Academic Achievement

Only the eighth grade students' scores were employed in this analysis, for the reason that scores for a previous administration (Spring, 1972) of the same test forms were available. Previous scores for seventh graders were obtained by means of a different test form from that employed during the January, 1973 testing. In the analysis of eighth grader scores, change scores were used; the investigators were primarily interested in net gain or loss in academic achievement between the two testing periods.

The analysis revealed that neither group had experienced any academic achievement loss between the two test periods: that is, all differences were positive. However, the analysis showed no differences between the two groups nor for either sex. The interaction similarly did not attain statistical significance. A total of 104 subjects was employed in this analysis.

Student: Personal Growth

Because of doubts concerning the reliability of results obtained during the first administration of the Bell Adjustment Inventory (Fall, 1972) and because of substantial absences of students during both of the testing sessions, it became necessary to combine groups in order to obtain a sufficient N for analysis. (Employing change scores for this analysis, which would have reflected changes in the student's personal growth and adjustment between the two testing sessions, would have been preferable.) In view of these constraints, only raw scores obtained at the second testing session were used. Grade level was eliminated as an independent variable, and the results were analyzed only for the grouping variable and the sex variable.

The results of the multivariate analysis of variance showed no significance for the grouping variable but did, as expected, show statistical significance for the sex variable. The interaction of these two variables did not attain statistical significance. A total of 124 subjects was employed for this analysis.

Student: Attitudes

Change scores, reflecting differences in raw scores between the two administrations of the attitude scale, were analyzed. A full subgrouping analysis was conducted, employing all eight subgroups of the study's population. Two of the independent variables demonstrated statistical significance in this analysis: the grouping variable and the item variable.

No practical significance can be attached to the results obtained with respect to the item variable, since the scale was designed in such a manner as to cause the twelve paired items to tap different areas of the subject's attitudes.

The obtained F for the grouping variable was 5.280 (degrees of freedom = 1/184), which is significant at the 5% level only. Examination of the mean change for each group shows a very slight positive gain for the OSCAR group and practically no change for the control group.

A total of 192 subjects was employed in this analysis.

Students: Absences

As of January 31, 1973, 95 days of the 1972-73 academic year had been completed. A record of absences for each student in both OSCAR and control group was obtained. Only students who had been enrolled in their respective programs for the full 95-day period were considered in an analysis of absences. Separate analyses of variance were conducted for seventh graders and eighth graders, but sex and group were maintained as independent variables. Neither analysis showed any significant differences in frequencies of absences for the two groups. Each of the two separate analyses included a total of 100 students.

Teachers: Classroom Behavior

Data obtained during the two observational (Flanders) periods were separately analyzed through analysis of variance. Results with respect to the grouping variable were not significant in either analysis. The first analysis included data regarding the classroom behavior of the four OSCAR and the four control teachers, while the second analysis was based upon data from only three teachers in each group.

Teachers: Attitudes

Results obtained from administration of the attitude scale to OSCAR and control teachers were separately analyzed for each administration. Statistical significance for the grouping variable was not demonstrated in the course of either analysis. Data obtained from all eight teachers were employed in both analyses.

Parents: Attitudes

Analysis of variance was applied to data derived from the first administration of the parents attitude scale. Statistical significance was not demonstrated by the grouping variable. This analysis included data provided by 216 parents.

## Discussion

It is clear from these results that as of the midpoint of this study no significant differences between the OSCAR and control groups have been identified--neither with respect to the students themselves, nor their teachers, nor their parents.

At least three logical possibilities can be framed to account for these results: (1) the dependent variables chosen for this study are inappropriate to an evaluation of open education; (2) the educational program pursued for the benefit of the OSCAR students is not truly an "open" education program; (3) an open education program does not differ in its effect from a traditional program.

The first of these possibilities has been extensively considered by the investigators. It well may be that the instruments employed in this study to date simply do not measure the types of outcomes which follow or which can be expected to follow from an OSCAR-type program. While these instruments will continue to be employed during the balance of this study, the investigators intend to add further behavioral measures which, hopefully, will reflect true behavioral differences between the two groups. As noted earlier, one of the principal difficulties of this study has been that of defining behavioral objectives of the OSCAR program and contrasting these with behavioral objectives of the traditional program.

With respect to the second factor which may possibly account for present failures to demonstrate differences between the programs, results of the Flanders analysis indicated that teachers in the two programs do not behave differently within the rather limited scope of the Flanders categories. The investigators intend to employ other measures of teacher behavior to permit more detailed comparison and contrast between behaviors exhibited by the two groups of teachers. Should these continue to be non-significant, more intensive training of the OSCAR teachers in the methodology of open education may be appropriate.

It is the opinion of the investigators that open education has yet to be subjected to rigorous analysis. While the concepts of open education appear at the verbal level to differ significantly from those of traditional education, it well may be that in terms of performance outcomes, these two methodologies do not truly differ.

The balance of the present academic year's study will be devoted to a continuation of widespread searching for appropriate dependent variables, particularly those which will reflect behavioral objectives. At the conclusion of this study, a discriminant function analysis will be conducted in an effort to differentiate between "successful" students and unsuccessful students in the two groups. It is hoped that the outcome of this analysis can be

tested through appropriate assignment of students to the OSCAR and control program during the following academic year; these students' progress will provide a test of the efficiency of the discriminant function. Clearly, considerable additional research based upon sharper delineation of educational objectives will be necessary before a definitive opinion can be rendered with regard to the efficacy and efficiency of open education.

A final report of this year's study will be prepared shortly after the end of the academic yaer and will be made available to those interested.